

# Protein frustratometer: a tool to localize energetic frustration in protein molecules

Michael Jenik<sup>1,2</sup>, R. Gonzalo Parra<sup>1</sup>, Leandro G. Radusky<sup>1</sup>, Adrian Turjanski<sup>2,3</sup>, Peter G. Wolynes<sup>4</sup> and Diego U. Ferreiro<sup>1,\*</sup>

<sup>1</sup>Protein Physiology Laboratory, Departamento de Química Biológica—CONICET, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Buenos Aires C1428EGA, <sup>2</sup>Structural Bioinformatics Laboratory, Departamento de Química Biológica, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Buenos Aires C1428EGA, <sup>3</sup>Departamento de Química Inorgánica, Analítica y Química Física/INQUIMAE, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Buenos Aires C1428EHA, Argentina and <sup>4</sup>Department of Chemistry, Rice University, Houston, TX 77251, USA

Received February 24, 2012; Revised April 21, 2012; Accepted April 29, 2012

## ABSTRACT

**The frustratometer is an energy landscape theory-inspired algorithm that aims at quantifying the location of frustration manifested in protein molecules. Frustration is a useful concept for gaining insight to the proteins biological behavior by analyzing how the energy is distributed in protein structures and how mutations or conformational changes shift the energetics. Sites of high local frustration often indicate biologically important regions involved in binding or allostery. In contrast, minimally frustrated linkages comprise a stable folding core of the molecule that is conserved in conformational changes. Here, we describe the implementation of these ideas in a webserver freely available at the National EMBNet node—Argentina, at URL: <http://lfp.qb.fcen.uba.ar/embnet/>.**

## INTRODUCTION

The energy landscape theory of protein folding is based on a statistical description of a protein's potential energy surface. Globally, the most realistic model of a protein is a minimally frustrated heteropolymer with a rugged funnel-like landscape biased toward the native state (1). This statistical description has been developed using tools from the statistical mechanics of disordered systems, polymers and phase transitions of finite systems. Natural proteins, as we observe them today, are highly evolved complex systems. Self-assembly of and mutual recognition between these polypeptides leading to reasonable well-defined structural ensembles is a fundamental concept in

the biology of macromolecules. The specificity of folding and binding is captured by the 'Principle of minimal frustration' (2). This principle states that the general energy of the protein decreases more than what may be expected by chance as the protein assumes conformations progressively more like the ground (native) state. In other words, there is a strong energetic bias toward the native basin that overcomes both the asperities of the landscape which stabilize kinetic traps and also ultimately the entropy of the chain. It has been shown that the structures of transition state ensemble (3,4), the folding rate variations (5), the existence of folding intermediates (6), dimerization mechanisms (7) and domain swapping events (8) are often well predicted in models where energetic frustration has been removed from the model landscape and topological information of the native state is the sole input. Still, inhomogeneity in the native contacts energetics, non-native interactions and the residual local frustration present in the native ensemble do contribute to the functional characteristics of proteins, 'molding' the roughness that underlies the detailed protein dynamics (9,10).

## Local frustration

The principle of minimal frustration does not rule out that some energetic frustration may be present in a folded protein. Moreover, the remaining frustration may not be random but evolved, facilitating motion of the protein around its native basin, and as such the residual frustration may be fundamental to protein function (9,10). Theoretical methods allow for spatially localizing and quantifying the energetic frustration present in native protein structures by developing a spatially local version of the global gap criterion formulation of the minimal frustration principle (11). This algorithm compares the

\*To whom correspondence should be addressed. Tel: +54 11 4576 3300; Fax: +54 11 4576 3342; Email: ferreiro@qb.fcen.uba.ar

contribution to the extra stabilization energy ascribed to a given pair of amino acids in the native protein to the statistics of the energies that would be found by placing different residues in the same native location or by creating a different environment for the interacting pair. If there is a sufficient additional stabilization for an individual native pair as normalized by the typical energy fluctuation (in accord with the global Z-score criterion for minimal frustration) the local interaction can be called 'minimally frustrated'. If the stabilization of the native pair lies in the middle of the distribution of alternatives, the interaction can be considered 'neutral'. On the other hand, if the native pair is sufficiently destabilizing compared with the other possibilities, the interaction is 'highly frustrated'. Details of the method and the energy functions can be found in references (11,12,13) and at the server documentation pages.

## MATERIALS AND METHODS

### Input files

The main input file contains a set of protein structure coordinates in the standard format of the Protein Data Bank (<http://www.rcsb.org>). Users can upload their own structure file or provide the four-letter code for existing PDB entries, in which case the server will automatically download the data. The coordinates are checked for overall formatting and if there is more than one amino acid chain in the model, the user is asked to specify the chains to process. A dialogue box and an interactive Jmol

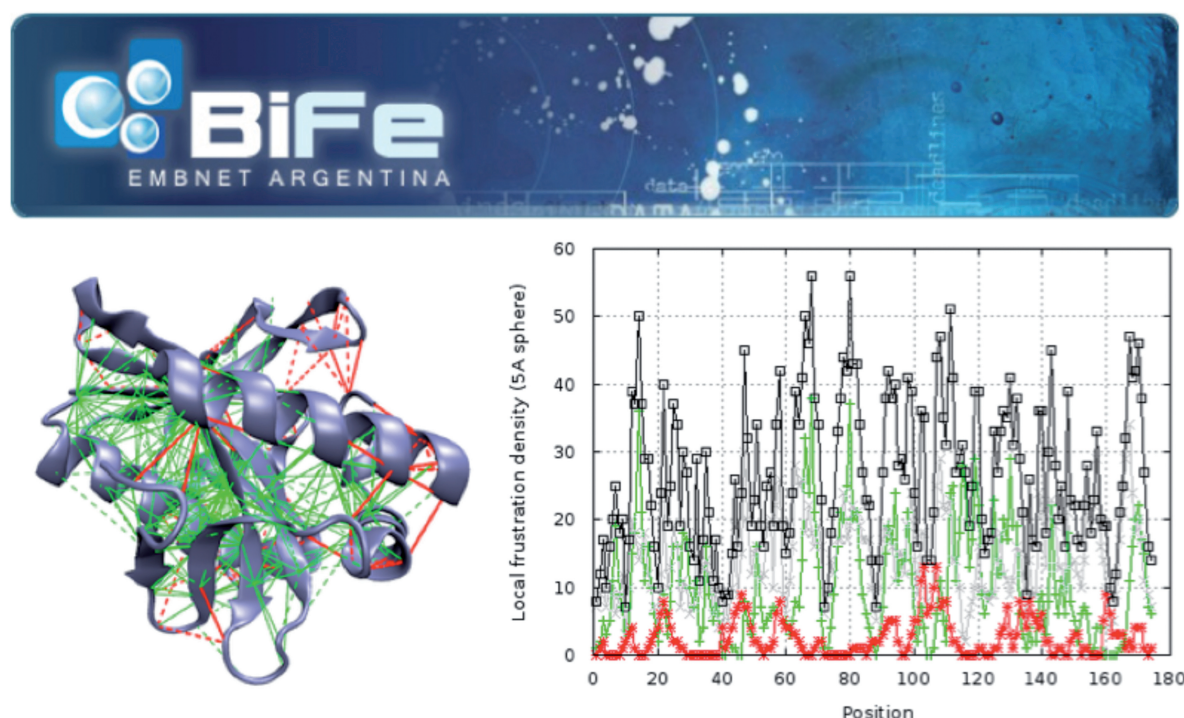
interface are provided to assist in this process. The user can specify as many chains as wanted. Jobs are automatically accepted for up to 1000 amino acid residues complexes. Users can optionally provide an e-mail address to receive a notification of job completion.

### Server calculations

The server automatically applies filters that remove hydrogens, heteroatoms and alternative conformations for residues. If the input file contains multiple models, only the first one is analyzed by default. The most common 20 amino acids are taken into account and if backbone or C-beta atoms are missing, they are automatically built into the file using an automodel option of Modeller suite (<http://salilab.org/modeller/>). The jobs are assigned a JobID, organized in a run queue and processed as computational resources become available. A typical run of a 200 residue protein takes about 5 minutes of CPU time and about 60 minutes for 500 residues complexes.

### Outputs

A results page is generated for each job. These pages can be accessed by following the link sent by e-mail (if provided) or by specifying the JobID. The server generates several projections of the local frustration calculations (Figure 1). An interactive Jmol applet facilitates inspection of the structures for which the minimally frustrated and highly frustrated contacts were identified. This information is also plotted as a contact map. Linear projections of local frustration distributions are also



**Figure 1.** Frustratometer server output. An example of the localized frustration and minimally frustrated networks in protein structures (pdb: 2FCK) Left: the protein backbone is displayed as blue ribbons, the direct inter-residue interactions with solid lines and the water-mediated interactions with dashed lines. Minimally frustrated interactions are shown in green, highly frustrated contacts in red, neutral contacts are not drawn. Right: Projection of local frustration distribution in amino acid sequence. The number of contacts within 5Å of the C-alpha of each residue is plotted, as classified according to their frustration index.

provided and can be enlarged in the same page. Results can be downloaded by following the link at the bottom of the same page. The download pack includes the input file as processed by the core algorithm, scripts to interactively visualize frustratographs in Jmol, PyMOL or VMD programs, together with the raw tables of the frustration index calculations and an explanatory README file. The results are accessible for 30 days and permanently deleted afterward.

## RESULTS

### Interpreting results

The frustration index measures how favorable a particular contact is relative to the set of all possible contacts in that location normalized using the variance of that distribution. For initial inspection, the server classifies the individual contacts as to their frustration index value. A contact is defined as 'minimally frustrated' (drawn green), if its native energy is at the lower end of the distribution of decoy energies, having a frustration index of 0.78 or higher magnitude (11), that is, the majority of (but by no means all!) other amino acid pairs in that position would be unfavorable. Conversely, a contact is defined as 'highly frustrated' (drawn red), if the native energy is at the other end of the distribution with a local frustration index lower than  $-1$ , that is, most other amino acid pairs at that location would be more favorable for folding than the native ones by more than one standard deviation of that distribution. If the native energy is in between these limits, the contact is defined as 'neutral' (drawn grey or not shown).

A frustration index may depend on the choice of parts in which the protein's whole energy is divided. It, therefore, becomes natural to divide the energy up in a way that is at least roughly comparable to what natural selection can do: examine the changes in energy on making mutations. The webserver provides two complementary ways for localizing frustration that differ in how the set of decoys is constructed:

#### *Mutational frustration (How favorable are the native residues relative to other residues in that location?)*

The decoy set is made randomizing the identities of the interacting amino acids, keeping all other interaction parameters at their native value. This scheme effectively evaluates every possible mutation of the amino acid pair that forms a particular contact in a robustly fixed structure. It is worth noting that the energy change on a residue pair mutation not only comes directly from the particular contact probed but also changes through interactions of each residue with other residues not in the mutated pair, and those contributions will also vary on mutation.

#### *Configurational frustration (How favorable are the native interactions between two residues relative to other interactions these residues can form in other compact structures?)*

This way of measuring frustration imagines that the residues are not only changed in identity but also may

be displaced in location. The energy variance thus reflects contributions to different energies of other compact conformations. In this calculation, the decoy set involves randomizing not just the identities and also the distance and densities of the interacting amino acids. This scheme effectively evaluates the native pair with respect to a set of structural decoys that might be encountered in the folding process.

### Case studies

A survey of nonredundant protein domains shows that natural protein domains are strongly crosslinked by minimally frustrated contact networks comprising about 40% of the total contacts (11). Only a minority ( $\sim 10\%$ ) of the native interactions are found to be 'highly frustrated', and these typically cluster at the protein surface (11). The remaining 50% are 'neutral' and are randomly distributed in the structure. The highly frustrated interactions that, in principle, might conflict with the robust folding of the domain seem to reflect evolutionary constraints other than folding and often correspond to physiologically relevant sites. A statistical survey shows that these sites do co-localize with regions involved in the formation of heterodimeric protein assemblies (11). A survey on the local frustration patterns of allosteric domains shows that the regions that reconfigure are often enriched in patches of highly frustrated interactions (12), consistent with the idea that these locally frustrated regions may 'crack' in these locations (14). On the other hand, the symmetry of multimeric protein assemblies allows near degeneracy by reconfiguring while maintaining minimally frustrated interactions (12). In addition, highly frustrated regions found in the native ensemble have been found to contribute to the stabilization of folding intermediates (15,16). In a similar spirit, PROSA web service displays Z-scores and energy plots that highlight potential conflicts spotted in protein structures (17).

### Concluding remarks

Natural protein domains must be sufficiently stable to fold but often also need to be locally unstable to function. The possibility of localizing and quantifying the energetic frustration present in protein molecules allows one to probe lower hierarchies of the energy landscape manifested as the exploration of the configurational substates defined by the local roughness. Molding of this roughness can have profound effects on the structural transitions and is thus likely to have functional consequences. Particular examples of frustratographs can be very interesting, but results should be taken carefully, as in some cases, the highly frustrated regions may not correspond to the known active sites. Performing statistical surveys of homologs, mutants, etc. are encouraged. The server is supported by a Documentation section, including a quick-start guide and walkthrough with screenshots. A gallery of frustratographs with examples is also hosted at the site, together with fully interactive outputs that are linked from the help pages. We also host a FAQs section and personalised support is provided via e-mail request.



## ACKNOWLEDGEMENTS

D.U.F. thanks D.J. Goldstein for early inspiration and question seeking.

## FUNDING

National Institutes of Health [R03 TW 008232]; National Institute of General Medical Sciences [R01 GM44557] and [P01 GM071862]. The content is solely the responsibility of the authors and does not necessarily represent the official views of National Institute of General Medical Sciences or the National Institutes of Health. D.U.F. is a researcher and R.G.P. holds a fellowship of Consejo Nacional de Investigaciones Científicas y Técnicas, CONICET, Argentina. Funding for open access charge: National Institutes of Health [R03 TW 008232].

*Conflict of interest statement.* None declared.

## REFERENCES

- Onuchic, J.N., Luthey-Schulten, Z. and Wolynes, P.G. (1997) Theory of protein folding: the energy landscape perspective. *Ann. Rev. Phys. Chem.*, **48**, 545–600.
- Bryngelson, J.D. and Wolynes, P.G. (1987) Spin glasses and the statistical mechanics of protein folding. *Proc. Natl Acad. Sci. USA*, **84**, 7524–7528.
- Clementi, C., Nymeyer, H. and Onuchic, J.N. (2000) Topological and energetic factors: what determines the structural details of the transition state ensemble and “en-route” intermediates for protein folding? An investigation for small globular proteins. *J. Mol. Biol.*, **298**, 937–953.
- Koga, N. and Takada, S. (2001) Roles of native topology and chain-length scaling in protein folding: a simulation study with a Go-like model. *J. Mol. Biol.*, **313**, 171–180.
- Chavez, L.L., Onuchic, J.N. and Clementi, C. (2004) Quantifying the roughness on the free energy landscape: entropic bottlenecks and protein folding rates. *J. Am. Chem. Soc.*, **126**, 8426–8432.
- Ferreiro, D.U., Cho, S.S., Komives, E.A. and Wolynes, P.G. (2005) The energy landscape of modular repeat proteins: topology determines folding mechanism in the ankyrin family. *J. Mol. Biol.*, **354**, 679–692.
- Levy, Y., Wolynes, P.G. and Onuchic, J.N. (2004) Protein topology determines binding mechanism. *Proc. Natl Acad. Sci. USA*, **101**, 511–516.
- Yang, S., Cho, S.S., Levy, Y., Cheung, M.S., Levine, H., Wolynes, P.G. and Onuchic, J.N. (2004) Domain swapping is a consequence of minimal frustration. *Proc. Natl Acad. Sci. USA*, **101**, 13786–13791.
- Frauenfelder, H., Sligar, S.G. and Wolynes, P.G. (1991) The energy landscapes and motions of proteins. *Science (New York, N.Y.)*, **254**, 1598–1603.
- Hegler, J.A., Weinkam, P. and Wolynes, P.G. (2008) The spectrum of biomolecular states and motions. *HFSP J.*, **2**, 307–313.
- Ferreiro, D.U., Hegler, J.A., Komives, E.A. and Wolynes, P.G. (2007) Localizing frustration in native proteins and protein assemblies. *Proc. Natl Acad. Sci. USA*, **104**, 19819–19824.
- Ferreiro, D.U., Hegler, J.A., Komives, E.A. and Wolynes, P.G. (2011) On the role of frustration in the energy landscapes of allosteric proteins. *Proc. Natl Acad. Sci. USA*, **108**, 3499–3503.
- Papoian, G.A., Ulander, J., Eastwood, M.P., Luthey-Schulten, Z. and Wolynes, P.G. (2004) Water in protein structure prediction. *Proc. Natl Acad. Sci. USA*, **101**, 3352–3357.
- Whitford, P.C., Onuchic, J.N. and Wolynes, P.G. (2008) Energy landscape along an enzymatic reaction trajectory: hinges or cracks? *HFSP J.*, **2**, 61–64.
- Ferreiro, D.U. and Wolynes, P.G. (2008) The capillarity picture and the kinetics of one-dimensional protein folding. *Proc. Natl Acad. Sci. USA*, **105**, 9853–9854.
- Sutto, L., Latzer, J., Hegler, J.A., Ferreiro, D.U. and Wolynes, P.G. (2007) Consequences of localized frustration for the folding mechanism of the IM7 protein. *Proc. Natl Acad. Sci. USA*, **104**, 19825–19830.
- Wiederstein, M. and Sippl, M.J. (2007) ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res.*, **35**, W407–W410.